

Gencore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 3, 2003, 21:11:56 ; Search time 407.991 Seconds
(without alignments) 16426.693 Million cell updates/sec

Title: US-10-027-000-1

Perfect score: 2976
Sequence: 1 ttatagtcgcttggtaaat.....aaaaaaaaaaaaaaaaaaaaa 2976

Scoring table:
IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	260.8	8.8	588	21	AAAF15000
2	184.4	6.2	3241	24	AAAF18442
3	179	6.0	2401	21	AAAF18442
4	179	6.0	2401	21	AAAF18442
5	177.4	6.0	1145	16	AAAF18442
6	175.2	5.9	2430	21	AAAF18442
7	175.2	5.9	2430	24	AAAF18442
8	175.2	5.9	12441	21	AAAF18442
9	175.2	5.9	13613	24	AAAF18442

10	167.2	5.6	13613	21	AAAF18442
11	161	5.4	2256	17	AAAF18442
12	138.6	4.7	50937	21	AAAF18442
13	135.8	4.6	2166	18	AAAF18442
14	135.8	4.6	2166	18	AAAF18442
15	111.6	3.8	4403765	22	AAAF18442
16	111.6	3.8	4411529	22	AAAF18442
17	103	3.5	985	24	AAAF18442
18	103	3.5	985	24	AAAF18442
19	101.4	3.4	16836	19	AAAF18442
20	98.6	3.3	3849	22	AAAF18442
21	93.4	3.1	2271	11	AAAF18442
22	91.2	3.1	12588	15	AAAF18442
23	85.8	2.9	65140	22	AAAF18442
24	85.4	2.9	125401	22	AAAF18442
25	85.4	2.9	125401	22	AAAF18442
26	84.4	2.8	390	13	AAAF18442
27	84.4	2.8	390	13	AAAF18442
28	84.4	2.8	390	13	AAAF18442
29	84.4	2.8	390	13	AAAF18442
30	83.8	2.8	3032	13	AAAF18442
31	83.8	2.8	3032	13	AAAF18442
32	83.8	2.8	3032	13	AAAF18442
33	80.4	2.7	1140	19	AAAF18442
34	80.4	2.7	1140	19	AAAF18442
35	80.4	2.7	1140	19	AAAF18442
36	78.6	2.6	3269	16	AAAF18442
37	78.6	2.6	3269	16	AAAF18442
38	78.2	2.6	1294	22	AAAF18442
39	78.2	2.6	1294	22	AAAF18442
40	78.2	2.6	1521	16	AAAF18442
41	78.2	2.6	1521	16	AAAF18442
42	78.2	2.6	1521	16	AAAF18442
43	78.2	2.6	1521	16	AAAF18442
44	78.2	2.6	1521	16	AAAF18442
45	78.2	2.6	1521	16	AAAF18442

ALIGNMENTS

RESULT 1	AAAF15000	standard; cDNA; 588 BP.
XX	AAAF15000;	
XX	13-MAR-2001 (first entry)	
DE	Trichoderma reesei EST SEQ ID NO:7523.	
XX	Multiple gene expression; filamentous fungal cell; EST;	
KW	expressed sequence tag; Fusarium venenatum; Aspergillus niger;	
KW	Aspergillus oryzae; Trichoderma reesei; identification; recombination;	
KW	culture condition; environmental stress; spore morphogenesis;	
KW	metabolic pathway engineering; catabolic pathway engineering; ss.	
XX	Trichoderma reesei.	
OS	WO200056762-A2.	
XX	28-SEP-2000.	
PD	22-MAR-2000; 2000WO-US07781.	
XX	22-MAR-1999; 99US-0273623.	
PF	(NOVO) NOVO NORDISK BIOTECH INC.	
XX	(NOVO) NOVO NORDISK AS.	
PA	Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;	
PI	WPI; 2000-594572/56.	
DR		

S. venezuelae deso
Chimeric thermost
Streptococcus olea
Thermotoga maritima
Thermotoga maritima
Mycobacterium tube
Mycobacterium tube
Oligonucleotide fo
Oligonucleotide fo
Acetobacter xyliu
S. chrysomallus ac
Kylase gene from
Entire amylose gen
Sequence encoding
Streptomyces nours
Streptomyces nours
Randomising oligon
PCR primer for 5'
Sequence containin
Oligo #7 for clon
Trichoderma reesei
DNA encoding a bet
Trichoderma reesei
Codon-optimised RA
Modified HIV prote
Corynebacterium gl
Streptomyces fradi
cDNA encoding aven
Streptomyces albid
Nucleotide sequenc
3-Hydroxysteroid o
3-Hydroxysteroid o
Streptomyces sp. s
3-Hydroxysteroid o
3-Hydroxysteroid o
Streptomyces sp. s

XX Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -

PS Claim 89; Page 3038-3039; 3161pp; English.

XX The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organization of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*
 CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
 CC all specifically claimed in the present invention.

XX Sequence 588 BP; 130 A; 167 C; 157 G; 109 T; 25 other;

Query Match 8.8%; Score 260.8; DB 21; Length 588;

Best Local Similarity 92.8%; Pred. No. 3.3e-33; Mismatches 20; Indels 2; Gaps 1;

Matches 283; Conservative 0; Mismatches 20; Indels 2; Gaps 1;

XX 2208 ATCCAGGCTGTGACGGCGGACGAGAGGCGGCACTCCATTCGCGAGCTGCTTTGGC 2267

DB 1 ATCCAGGCTGTGACGGCGGAGGAGGAGGCGGCACTCCATTCGCGAGCTGCTTTGGC 60

XX 2268 GACTACACACCCCTGGGCAACCTGCTCCCTCCCAAGCCGCTGCGAGCAACCC 2327

DB 61 GACTACACACCCCTGGGCAACCTGCTCCCTCCCAAGCCGCTGCGAGCAACCC 120

XX 2328 GCGTTTCAACTTCCGACCGGAGCGGCGGCGGCACTGTAAGCGGAGAGTACAGTC 2387

DB 121 GCGTTTCAACTTCCGACCGGAGCGGCGGCGGCACTGTAAGCGGAGAGTACAGTC 180

XX 2388 GGGTACAGGTACTACAGAGTTGGCGGACAGAGAGTCAATTTCCCTTTGGCCAGGCGCTG 2447

DB 181 GGGTACAGGTACTACAGAGTTGGCGGACAGAGAGTCAATTTCCCTTTGGCCAGGCGCTG 240

XX 2448 TCCACACCACTTTT--GCTTTTCAATCTCTCGTGTCTACAGAGAGCGGCAAGCTGA 2505

DB 241 TCCACACCACTTTTGGCTTTTTCATCTCTTCCGGGCTTAAACAGAGCGGCAAG 300

XX 2506 GCGTG 2510

DB 301 GCTTG 305

RESULT 2

XX AAS18442

XX AAS18442 standard; DNA; 3241 BP.

XX AAS18442;

XX 12-MAR-2002 (first entry)

XX Contlig 93 DNA encoding *S. narbonensis* polyketide synthase.

KW Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
 KW erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethylmycin;
 KW agriculture; ds.

XX Streptomyces narbonensis.

XX US6303767-B1.

XX 16-OCT-2001.

XX 05-NOV-1999; 99US-0434288.

XX 05-NOV-1998; 98US-107093P.

XX 27-MAY-1999; 99US-0320878.

XX (KOSA-) KOSAN BIOSCIENCES INC.

XX Betlach MC, McDaniel R;

XX WPI; 2002-065495/09.

XX Nucleic acids encoding narbonolide polyketide synthases from
 PT Streptomyces narbonensis, useful for the recombinant production of
 PT polyketides, e.g. narbomycin -

XX Claim 1; Column 20-22; 24pp; English.

XX The present invention relates to recombinant DNA vectors (cosmids).
 CC that encode for the narbonolide polyketide synthase (PKS) enzyme and
 CC various narbomycin modification enzymes from Streptomyces narbonensis.
 CC The recombinant DNA vectors can be used to produce recombinant ketide
 CC synthases and a variety of different polyketides (e.g. erythromycin,
 CC rapamycin, tylosin, narbomycin, picromycin, methylmycin and
 CC neomethylmycin) for use in agriculture, medicine and health. The
 CC recombinant vectors may be used to produce polyketides in relatively
 CC high yields. AAS18432-AAS18443 represent contig DNA sequences that
 CC encode for *S. narbonensis* PKS enzymes.

XX Sequence 3241 BP; 520 A; 1174 C; 1125 G; 422 T; 0 other;

Query Match 6.2%; Score 184.4; DB 24; Length 3241;

Best Local Similarity 48.2%; Pred. No. 6.8e-21; Mismatches 956; Indels 150; Gaps 12;

Matches 1028; Conservative 0; Mismatches 956; Indels 150; Gaps 12;

XX 544 CGCTCGTTCACATCTCAACCAAACTCTCTGAAAGAGGCAAGTATGATGAGGCAAG 603

DB 1151 CCTGCGCAGACACTTTCAGACACTCCATGCGGAGGCAAGGAGGATGAGGCGCGG 1210

XX 604 AGGCATTCGCTAAGATGGGATGATGATGATGATGATGATGATGATGATGATGATG 663

DB 1211 ACAGAGCGCGGCGGCGGCGGAGGAGGATGATGATGATGATGATGATGATGATGATG 1270

XX 664 CTCTCGTGTAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 723

DB 1271 CACAGGCGGCGGAGGAGTACAGAGCTTCAAGGAGAGGAGGAGGAGGAGGAGGAGG 1330

XX 724 CTGCGGCTCTCATCGGCGGCAATTCAGAGCAGTGAAGTGAAGTGAAGTGAAGTGA 783

DB 1331 CGGTGCGGCGGCAATTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1390

XX 784 TGTGCAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 843

DB 1391 CGGCGCAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1450

XX 844 TCCGTGAAATCTACAGCACTCCGTTCCAGATTGCTGTGAGAGACTCCAGGCGGCTGCT 903

DB 1451 TCCGGAGATGAGTTTCCGGGCTTCAG--GCTTCTCGAAGGCGGCGGCGCTCTCT 1507

XX 904 TCATGAGGCGGTACATGAGCATATGCGGTGCTGACGAGAGAACCTTAATATCTTG 963

DB 1508 TCATGAGGCGGTACATGAGCATATGCGGTGCTGACGAGAGAACCTTAATATCTTG 1567

XX 964 ATGGATGCTTGGAAGAAGATGGGCTTGAGATGAGGCTTAATCATGAGCATGATGAGCA 1023

Db 1568 ACAAGCTGCTGGCAGCAGCGAGGCTTCCAGGCTGGGTATGTCACAGTGGCTGGCA 1627
 Qy 1024 CATACAGTACACAGAGAGCCGTTGTGGCAGGCTCGACCTGAGATGGCCGAGCTCCAC 1083
 Db 1628 C---CCCGGGCAGAGCAGCATCACCAAGGCGCTCGACAGAGAT---GGGCTTCAG 1680
 Qy 1084 GCTTCGAGGAGAAACATCAAGTTCAAGTCTCCAGCGGAAAGCCCTTATCCACGTCA 1143
 Db 1681 CTCCTCCGGGACATCCCGCGGCGAGCCCTCGCGCCCAATTTCTCGGTACGGG 1740
 Qy 1144 TTGACAGAGGGGTAGAGAGTTCTTCACTTGTCAAGAGTGTCTGCTCCGAGTGA 1203
 Db 1741 CTGA-----AGCAGCCGTCTCTGAAGGACAGCGTCCCGAGGGCG 1780
 Qy 1204 CGGAGAACGCGCCCGAGAGCAGCTGTACACAGACCCCGAAAGCGAGCTCTCCGGA 1263
 Db 1781 CCGTACGCGGTGCGGAGAGCGCATCTGTCACAGATGAGCAAGTTCCGGTCTCTCG 1840
 Qy 1264 AGGTTGGCAACGAGGATCTGCTGCTGAAGAGAGAAACAGCTTCCGCTTGAGCA 1323
 Db 1841 CGACTCGGCGGCCCGCCCGAGCGTACACAGGCGGCGCCAGGCGGTGTCGCGAAG 1900
 Qy 1324 AGAAGAGAGAGAGCTGATTTGTGCGCCCGCAGCGCAGAGCGCAGATACAGCGGAG 1383
 Db 1901 TCGCGAGAGAGCGCGCGGTGCTGCTGCGACAGAGGCGCAGGCGCTGCGCGGTG 1960
 Qy 1384 GCTCTCCGCACTACAGGCTTACAGCAGTCACTCCCTTTGACGCGCTCAGCAAGCAG 1443
 Db 1961 ACCCGCGAGAGAGCATGCGCTCATCGCGCGCGCGCTGAGC---CCCAAGGTACAGCG 2019
 Qy 1444 TCAGAGAGCGCGCATCTGATACCGCTGCGGCGCTACACACACCGTCTTCCATTTAG 1503
 Db 2020 CTGGGAGCGCGCCCAAGTCTGCTCGGAGCTGCGGCGCGCGCTGACACATCAAGGCG 2079
 Qy 1504 AGCAGTGCCTACAGCGCGCGAGCGGCTCGGCGAGTGGCTGAGAGGCTTTCACAGAG 1563
 Db 2080 CGCGCGGCGCGGCTGAGAGTACAGTACAGAGAGGCTGAGAGAGCTTTCGCGAGCG 2139
 Qy 1564 CTGGTACCCCTAACCGCGACATTTGACAGAGCTCTTTCACACAGAGAGGATCCAGC 1623
 Db 2140 ATCCCGCGCGGAGCTCAGC-----CCGCGGTTCAACC 2173
 Qy 1624 TGTGAGCTACTACACCCCAAGCGGAGAGACGTGTACGCGGAGATGAGGCGACGT 1683
 Db 2174 AGGCGCCACAGCTGAGAGCGGAGAGCGGCGGCGCTGTACAGAGGAGCGTACGCTG 2233
 Qy 1684 ACACCGCGAGAGAGTGTACAGAGCTGCGCTGCTGCTGCGGAGCGGAGAGG 1743
 Db 2234 CCGCGAGCGGAGTACCGCATCTCGGTCAAGGCGCACCGGCTACGCGAGCGTCCAG- 2292
 Qy 1744 CGTACGTACAGAGAGCTGTGTGACAGACCGCAAGAGTCCCGCGGATGGCT 1803
 Db 2293 -----CTCGGAGCCACA 2305
 Qy 1804 TCTTCGCTCCGCCACCGCGAGAGAGCGGCGCATCATCTGTCAGAGGCGACAGCT 1863
 Db 2306 CCATCGAGGCGGTACAGTGTACGAGAGTGAAGCGCGCTCTCAAGCTGACCAAG 2365
 Qy 1864 ACAAGTTCAAGATGAGTGTGCTCGGACCCACCTACACCTTCAGAGGCGAGACCATG 1923
 Db 2366 GACGCGACAGCT-----CAGCATCTCGGGTTGCGCATGA 2401
 Qy 1924 TCCCGGCGCAGGCTCTCGGCTCGGCTGCAAGAGTCAATTTGAGAGCAGCGGAA 1983
 Db 2402 GCGCGACCGCGCTCTCTGAGAGTGGGTGAGCGCGGAGCGAGCGGAGCGGAGCA 2461
 Qy 1984 TCGAAGATCGTCCGCTCGCAGAGAGAGCAGCAGAGTCACTATCTGCGCGGCTTGA 2043
 Db 2462 TCGGAGAGGCGGTGAGTGTGCGCGGAGAGCGCGGTACGCGCATCTGTTGGC----- 2514
 Qy 2044 ACGCCGACTGGAGAGCGAGGCGCGAGCGCGGAGAGCATGAAGCTTCCCGGCGTGTG 2103

Db 2515 --TACAGAGAGCGAGCGAGGCGCTGACACCTCCGAACCTGTGCTCCGAGTACGAGG 2572
 Qy 2104 ACCAGCTATTTGCGGAGTGGCGCGCGGAGAACCCAAACACCGTGTGATGAGAGG 2163
 Db 2573 ACAAGCTGATTCGCGGCTGCGCACAGCGAGCCGAACACAGATGTGTCTTCAACCG 2632
 Qy 2164 GCACCCCGAGAGATGCTCTGCTGAGAGCGGAGCGCGCGCTGATCCAGGCTGTACG 2223
 Db 2633 GTTCGTGCTGTATGCGGTGCTGTCAGAGACCGCGGCGGTCTGTGACATGTTGTAAC 2692
 Qy 2224 GCGGACAGAGAGCGGCACTTCCAGCTGTCGCTTTGAGGAGTACACCCCTCGG 2283
 Db 2693 CGGCGAGGCGGCGCGCGAGCGAGCGCGCTGCTTACGAGTGAAGAGCG 2752
 Qy 2284 GCAGCTGTCTCTCACTTC-----CCAGCGCTGACAGAGACACCGCGCTTCTCA 2337
 Db 2753 GCAGCTACGAGAGCTTCCCGCGCGCGCGGAGAACAGAGACCGCTGCGCGAGCCGA 2812
 Qy 2338 ACTTCCGAGCGAGCGCGGCGCAGCGT--GTACGCGAGAGAGCTTACGTGGTACA 2394
 Db 2813 ACCGCTACCGGCGGTGACAAACAGCAGCTACAGAGGAGCATCACCTCGGTTACC 2872
 Qy 2395 GGTACTACGAGTTTGGCAGCAAGAGCTCATTTCCCTTTGGCAGGCGCTGTCTACA 2454
 Db 2873 GCTGTTGACAGAGAGAGAGCTCAAGCGCGTGTCTCTTGGGAGCGCTGTGTACA 2932
 Qy 2455 CCACCTTTGCTTTTCAATCTCTCGTGTCTACAGAGAG---GGCAAGTGTAGCGTGT 2511
 Db 2933 CTCTGTTACGAGAGAGCGCGCGCGAGCTGTGAGAGAGCTTCCAGGCGCGCTGAAG 2992
 Qy 2512 CCCTCTCGTGAAGAGACACCGGCTCGTCCCGCGGCGCAGAGTGGCCCAAGTCTACGTA 2571
 Db 2993 CGGTACGAGTGGCGAACAGCGGCGGAGCGCGGCGGCGAGAGGTCTCCAGGCGTATCTG 3052
 Qy 2572 AGCCCTCCAGAGCGGCGCAAGATTACCGCGCGCTGACAGAGTCAAGGCTTTCGAAAG 2631
 Db 3053 GCGGAGCGCGAGAGTGTACGCGCTCCGAGCGGAGAGAAAGCTGTGGGCTACAGAGG 3112
 Qy 2632 TCGAAGTGTAGCGCGCGGAGAGAGAGCGGTGAC 2665
 Db 3113 TCGCGCTGCGGCGGCGAGTGTAGAGAGGTGAC 3146
 Db
 RESULT 3
 ID AAT75634
 AAAT75634 standard; DNA; 2401 BP.
 XX
 AC AAT75634:
 XX
 DT 22-JAN-2001 (first entry)
 XX
 DE Nucleotide sequence of ORF11 which encodes 1-beta glucosidase.
 XX
 KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
 KW antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;
 KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
 XX picromycin biosynthesis; ss.
 OS Streptomyces venezuelae.
 XX
 PN US6117659-A.
 XX
 PD 12-SEP-2000.
 XX
 PF 27-MAY-1999; 99US-0320878.
 XX
 PR 28-MAY-1998; 98US-0087080.
 PR 22-SEP-1998; 98US-0100880.
 PR 08-FEB-1999; 99US-0119139.
 PR 20-MAY-1999; 99US-0134990.
 PR 30-APR-1997; 97US-0846247.
 PR 06-MAY-1998; 98US-0073538.
 PR 28-AUG-1998; 98US-0141908.

XX (KOSA-) KOSAN BIOSCIENCES INC.

PA Ashley G, Betlach MC, Betlach M, Tang L, Medaniel R;

XX WPI: 2000-610844/58.

XX New recombinant pick hydroxylase gene of *Streptomyces venezuelae* useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value

XX Disclosure: Columns 39-40; 117pp; English.

XX The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.

XX Sequence 2401 BP; 405 A; 859 C; 828 G; 308 T; 1 other;

Query Match 6.0%; Score 179; DB 21; Length 2401;

Best Local Similarity 47.7%; Pred. No. 4.9e-20;

Matches 1018; Conservatively 0; Mismatches 966; Indels 150; Gaps 11;

QY 544 CGCTCGGTTCCATTCACCAACCTCTGTCGAAGAGCGAGTAAGTATGAGGCAAG 603
DB 255 CCGTGGCCAGACCTTCGACGACCAATGCGGACAGCTACGAGGCAAGTCAATGGCCGCG 314
QY 604 AGGCATTCGTAAGAGTGCATGTGATCTCGGCGGACATCAACATGCAAGCTCCG 663
DB 315 ACGGTGCGGCGCTCAACAGGACATGCTCTGGGCGCGGATGATACAAACATCCGSGTGC 374
QY 664 CTCTCGGTGACGTGGCTTCGATGATGATGATGATGATGATGATGATGATGATGATG 723
DB 375 CGACGCGGCGGCGGACGACCTTCAGCGAGGACCCCTCTCTCTCGGCGACCG 434
QY 724 CTGGGCTCTCATCCGGGCGGATTCAGAGCACTGAGTGCAGGCTACGATCAAGACTTTT 783
DB 435 CGGTGCGGCGGACATCAAGGCGATCCAGGCTGCGGCTGATGATGACCAAGGCACTTTCG 494
QY 784 TGTGCAATGATCAGAGGACAGGCGCATGATGATGATGATGATGATGATGATGATGATG 843
DB 495 CGGCGCAACACAGGAGAGAACCGCTTTCGCGTGAAGCGCAATGTCGAGAGAGAGAGCG 554
QY 844 TCGGTGAATTCAGCACTCCGCTTCAGATGCTGTCGAGAGACTCCAGCGGGTGGCT 903
DB 555 TCCCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 611
QY 904 TCATGAGGCGGTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 963
DB 612 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 671
QY 964 ATGGAGTCTTCGAAAGGATGAGGTTGGAGTGCCTAATCATGAGGAGTGTGAGGCA 1023
DB 672 ACAAGGTGCTGCGGACGACATGAGGCTTTCAGAGGCTGAGTGTGATGATGATGATGATG 731
QY 1024 CATGACATACAGAGAGCGGTTGTGCGAGGCTTCGACCTCGAGATGCCGAGCTCCAC 1083
DB 732 C---CCGCGGCGACGAGCATCAACAAAGGCGCTCGACAGAGATG-----775
QY 1084 GCTTCGAGAGGAAAGCTCAAGTTCACAGTTCGAAAGGAAAGCCCTTATCCAGTCA 1143

DB 776 -----GGCTGTGAGACTCCCGGCGAGCTCCCGAAGGCGAGCCCTCGCGCGCA 827
QY 1144 TTGACAGAGGCGCTAGGAGATGTTCTTCATGTCGTAAGAGTGTGCTCGGAGTCA 1203
DB 828 AGTTCTTTCGCGA---GGCGCTAAGAGCGGCTCTCTGAAAGGCGACAGGCTCCGAGGCG 884
QY 1204 CGGAGAAAGGCGCGCGGAGAGGAGTCAACAAACACCCCGAAAGGCGAGCTCTCTCGGA 1263
DB 885 CCGTGAAGCGGCTGCGGAGCGAGATGCTGCGGCAAGATGAGAGAGTGTGCTCTCTCGG 944
QY 1264 AGTTGGCAAGAGGCGATGCTGCTGTAAGAGAGAGAAACAGTTGCTCCCTTGAGCA 1323
DB 945 CCACTCCGCGCGCGCGCGCGCGAGCAAGGCGGCTCCAGCGGCTCCCGCAAG 1004
QY 1324 AGAAGAAAGAGAGCGATGATGTCGGCCCAAGCGGCAAGAGGCGACATACAGCGGCGAG 1383
DB 1005 TCGCGGAAAGCGCGCGGCTCTCTGCGCAAGAGGCGGCGGCTGCGGCTGCGGCTG 1064
QY 1384 GCTTGTCCGCACTCAGAGGCTCTACTAGCAAGTCACTCCCTTTGACGCGCTCAGCAAGCAG 1443
DB 1065 ACGCGGCAAGAGCATGCGGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1123
QY 1444 TCGAGAGCGCGCGCATGCTGACACCGTGGCGCTTACACACCGTTCTCCATTTAGCG 1503
DB 1124 CTGGGAGCG 1183
QY 1504 AGCAGTGCCTCAGCG 1563
DB 1184 CGCGCGGCTGCGGCTGCG 1243
QY 1564 CTGATCCCTTAAACCG 1623
DB 1244 ATCCGCGCGGGAAGCTCAGC-----CCGCGTTCAAC 1277
QY 1624 TGTGGAATCTACTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1683
DB 1278 AGGCGCACAGCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1337
QY 1684 ACACCGCGCGCGAGGAGCTGACCTGACCTGAGCTGCGCGCGCGCGCGCGCGCGCG 1743
DB 1338 CGCGCGCGCGCGAGATGCGGATGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1396
QY 1744 CGTACGTAGACGACGAGTCTGCTGACACGCGCACCAAGAGGTCCCGGCGATGCT 1803
DB 1397 -----CTGCGCPACCA 1409
QY 1804 TCTTGGCTCCCGACCGCGCGAGAGAGCGCGCGCATCATCTCTCAAGGCGCAACAGT 1863
DB 1410 CCAATGAGGCGGTCAGTCTACGCGCAAGTGAAGCGCGCGCTCTCAAGCTGACCAAG 1469
QY 1864 ACAGTTCAAGATCGAGTTGCGCTCCGCGCACCACTACACCTCTCAAGGCGCAACCATCG 1923
DB 1470 GCAAGCGCAAGCT-----CAGCATCTCGGCTTCGAGTGA 1505
QY 1924 TCCCGGCGACGCGCTCCCTCGCGCTGCGCGCTGCAAGGTCTTGAAGCAACAGCGGAAA 1983
DB 1506 GTGCGCACCGCGCTCTCCGAGCTGAGGCTGCGGTGAGTACCGCGCGCGCGCGAGCA 1565
QY 1984 TCGAAAGTCTCGCTTA 2043
DB 1566 TCGGAAAGGCGGTGAAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1617
QY 2044 ACGCGCGCTGAGAGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2103
DB 1618 -CTAGCAGCAGCGCACCGAGGCGCTGCGACCGTTCGATCTGCTGCGGCTGACGAGG 1676
QY 2104 ACCAGTCAATTCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2163
DB 1677 ACAAGCTGATCTCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1736
QY 2164 GCAAGCGCGAGAGATGCGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2223
DB 1737 GTTCGTGCGTGTGATGCGCGGCTGCTGCAAGACCGCGCGGCTCTGGAATGTGTGAC 1796

QY 2224 GCGGCAACGAGACGGGCACTTCATTCGCGAGCTGCTTTGGCGATACACCCCTCGG 2283
 DB 1797 CCGGCAAGCGGGGCGGAGCGCCACCGCCGCTGCTACGGGTACGCTCAACCGGAGCG 1856
 QY 2284 GCAAGCTGTCCTCAGCTTCCCGC-----AAGCGCTGCAAGCAACCCCGGTTTC 2334
 DB 1857 GCAAGCTGTCAGGAGCTTCCCGGCGCGCGAAGACGACGCGGCTCGCGGCGACCGGA 1916
 QY 2335 TCACCTTCGCGACCGGCGCGCGCGCGCGCTGACGGGAGGAGGCTGACGCGGCTACA 2394
 DB 1917 CCAGCTACCGCGGCGCTGACACGACGACGACGACGACGACGACGACGACGACGACG 1976
 QY 2395 GGTACGACGATTCGCT 2454
 DB 1977 GCTGCTTCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2036
 QY 2455 CCAGCTTTCCTTTCATCTCTCCGCTGCTACAGAGAC---GGCAAGCTGAGCGGT 2511
 DB 2037 CCTGCTTCAGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2096
 QY 2512 CCCCTTCGTAAGAACACCGCGCTCCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2571
 DB 2097 CGGTACGCTTCGCT 2156
 QY 2572 AGCCCTTCGCT 2631
 DB 2157 GTGCGAGCT 2216
 QY 2632 TCGAAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2665
 DB 2217 TCTGCTGCT 2250

RESULT 4

AA256002 standard; DNA: 2401 BP.

AA256002;

23-MAR-2000 (first entry)

Contig 001 from cosmid PKOS023-27 from Streptomyces venezuelae.

Narboxonolide polyketide synthase; PKS; cosmid PKOS023-27; contig 001;

ketolide; 1 beta glucosidase; antibiotic production; narboxonolide;

Streptomyces venezuelae.

Key Location/Qualifiers
 CDS 80..2389
 /tag= a
 /product= 1_beta-glucosidase

MO9961599-A2.

02-DEC-1999.

27-MAY-1999; 99WO-US11814.

28-MAY-1998; 98US-0087080.

28-SEP-1998; 98US-0141908.

08-FEB-1999; 99US-0119139.

(KOSA-) KOSAN BIOSCIENCES INC.

Ashley G, Belach MC, Belach M, McDaniel R, Tang L;

WPI; 2000-072618/06.

P-PSDB; AAY67212.

PT New recombinant DNA encoding a domain of narboxonolide polyketide
 synthase, for production of ketolide antibiotics -
 PS Disclosure; Page 34-35; 98pp; English.
 XX

This is contig 001 from the recombinant cosmid PKOS023-27 DNA sequence
 (see AA256001) which contains a Streptomyces venezuelae DNA insert. The
 cosmid contains open reading frames which encode the various modules of
 the narboxonolide polyketide synthase (PKS). The invention relates to
 recombinant DNA containing a coding sequence for a narboxonolide PKS.
 Polyketides are compounds synthesised from 2-carbon units through a
 series of condensations and subsequent modifications including picrotoxin.
 The narboxonolide PKS consists of a loading module, six extender modules,
 and two thioester domains. Four proteins make up the narboxonolide PKS
 (PICAI, PICAILI, PICAIIL and PICAIIV). PICAI includes extender modules 3 and 4,
 PICAILI includes extender module 5 and PICAIIV includes extender module 6
 and a type II thioesterase domain. The second type II thioesterase
 domain is found on the PICB protein. The nucleotide sequences encoding
 all of these proteins can be isolated in recombinant form from the
 recombinant cosmid PKOS023-27. Narboxonolide is desosaminylated in S.
 venezuelae to yield narboxonolide, and the desosaminyl transferase enzyme is
 required for this conversion. The desosaminyl transferase gene is
 also found in cosmid PKOS023-27. The recombinant DNA of the invention is
 used to express, in transformed cells, narboxonolide (or its derivatives)
 or other ketolides (particularly hybrids), which may then be converted
 (e.g. by other enzymes recombinantly expressed in the same hosts) to
 polyketide antibiotics or their intermediates. The antibiotics are useful
 in human or veterinary medicine.

Sequence 2401 BP; 405 A; 859 C; 828 G; 308 T; 1 other:

Query Match 6.0%; Score 179; DB 21; Length 2401;
 Best Local Similarity 47.7%; Pred. No. 4,9e-20;
 Matches 1018; Conservative 0; Mismatches 966; Indels 150; Gaps 11;

QY 544 CGCTCGGTTCCATTCACCAAACTCTGCTCGAAGAGGAGGAGTATGATGAGCAAG 603
 DB 255 CCTGTGCGCGGAGCTTCACGACACCATGCGCGCGCGCGCGCGCGCGCGCGCGCG 314
 QY 604 AGCGCATGCGTAAGAGTCCGATGATCTCTGCGCGCGCGCGCGCGCGCGCGCGCG 663
 DB 315 AGCGTGGCGCGCTACACGACGATGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 374
 QY 664 CTCTCGGTGAGAGTGGCTTCGATGATGATGATGATGATGATGATGATGATGATG 723
 DB 375 CGCAGCGCGCGCGGAACTACGACCTTCAGCGAGGAGCGCGCGCGCGCGCGCGCG 434
 QY 724 CTGCGGCTCTCATTCG 783
 DB 435 CGGTGCG 494
 QY 784 TGTGCAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 843
 DB 495 CGGCGCAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 554
 QY 844 TCCGTAATTCAGACCTCCGCTTCGATGATGATGATGATGATGATGATGATGATG 903
 DB 555 TCCGCGAGATGAGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 611
 QY 904 TCATGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 963
 DB 612 TCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 671
 QY 964 ATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1023
 DB 672 ACAAGCTGCTGCG 731
 QY 1024 CATACAGTACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1083
 DB 732 C---CCGGGCG 775

OY 1084 GCTTCGAGAGAAACATCAAGTTCACGCTCCACGAGAAAGCCCTTATCCAGCA 1143
 Db 776 -----GGCGTCGAGCTCCCGGCGAGCTCCGAGAGGAGGCCCTCGCGCGGCCA 827
 OY 1144 TTGACAGAGGGGTAGGGAAGTTCTTCAGTTCTGAATAAGTGTGCTCCGAGTGA 1203
 Db 828 AGTTCTTGGGCA---GGCGTGAAGAGCGGCTCTGAAGGAGCAGGTTCCCGAGCGG 884
 OY 1204 CGAGAAAGCGCCCGAGAGGACTCTCAACACACCCCGAGAGGAGGAGCTCTCCGGA 1263
 Db 885 CGGTGACGGGTGCGGAGGAGGATGTCGGGCAAGATGGAATTTGGTCTGCTCTCG 944
 OY 1264 AGTTGGCAAGAGGAGGATCTGCTGCTGAAGAAAGAGAACACGTTCTGCTTGAACA 1323
 Db 945 CCATCTCGGCGCGCGCGCCGAGCGGAGAGGCGGTGCTCCAGCGGTCTCCGCAAG 1004
 OY 1324 AGAAGAAAGAGAGCTGATTTGCGGCGCCACAGCGCAAGAGGAGCAATACAGGCGAG 1383
 Db 1005 TCGCGGAGAGAGCGCGGTGCTCTGCGCAACGAGGCGCAGGCTGCGCTGCGGCG 1064
 OY 1384 GCTTGGCGGCACTCAGGCGCTTACTACGAGTCTACCTCTTGAAGGCGCTAGCAAG 1443
 Db 1065 AGCGCGGAGAGAGATCGCGGTCTATCGCCGAGCGCGCTGAGC-CCAGAGTCAAGCG 1123
 OY 1444 TCGAGACGCGCGCATCTGACAGCGTGGCGCTTACACAGGCTTCTCCATTTAGCG 1503
 Db 1124 CTGGGCAAGCGCGCATCTGCTCCGAGCTCGCGGCGCGCGCTCTGACACATCAAGGCC 1183
 OY 1504 AGCAGTCTTACCGCGCGAGCGGCTCGGCGATGCGCTGAGAGGCTTCAAGAGAGCGCC 1563
 Db 1184 CGCGGAGGTGCGGCTGCGAGCGGTGACAGAGCGGAGGAGAGACCTTCGAGAGCGAG 1243
 OY 1564 CTGAGACCTTACCGCGCGAGCATTTAGAGAGCTTTTCAACAAGAGAGAGAGCAGC 1623
 Db 1244 ATCCCGGCGGAGACCTCAGC-----CCGCGCTTCAAC 1277
 OY 1624 TGTGAGCTACTTACACACCGGAGCGGAGAGAGAGAGCGGAGTGTGAGAGGAGCGT 1683
 Db 1278 AGGGCGACAGCTGAGCGCGGAGAGCGGCGGCGGTGATGAGAGCGGACGCTGAGCGTGC 1337
 OY 1684 ACAGCGCGAGAGAGAGTGTGAGTGTGAGTGTGCGCTGCTGCTGCGGAGCGGAGAG 1743
 Db 1338 CGCGGAGCGGAGTGTGAGTGTGAGTGTGCGCTGCTGCTGCGGAGCGGAGAG-1396
 OY 1744 GATAGTGAAGAGAGAGTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGTGTGCT 1803
 Db 1397 -----CTCGGAGAGCACA 1409
 OY 1804 TCTTCGCTCGGCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1863
 Db 1410 CATTCGAGCGGTGAGGTCTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1469
 OY 1864 ACAAGTTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 1923
 Db 1470 GCAAGCACAAGT-----CAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 1505
 OY 1924 TCCCGGCGAGCGCTCTCGCGGTGCGGCGGTGCAAGTCTTGAAGAGAGAGAGAGAG 1983
 Db 1506 GTGCGACACCGCGCTCTCGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1565
 OY 1984 TCGAAGTTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTC 2043
 Db 1566 TCGCGAAGCGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1617
 OY 2044 AGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2103
 Db 1618 -CTAG 1676
 OY 2104 ACAGAGTCTTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2163
 Db 1677 ACAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1736
 OY 2164 GCACCGCGAGAGAGATGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2223

Db 1737 GTTCGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1796
 OY 2224 GCGGCAACAGAGAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2283
 Db 1797 CGGCGAG 1856
 OY 2284 GCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2334
 Db 1857 GCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1916
 OY 2335 TCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2394
 Db 1917 CCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1976
 OY 2395 GATGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2454
 Db 1977 GCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2036
 OY 2455 CCAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2511
 Db 2037 CCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2096
 OY 2512 CCTCTCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2571
 Db 2097 CGGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2156
 OY 2572 AGCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2631
 Db 2157 GTGCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2696
 OY 2632 TCGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 2696
 Db 2217 TCTGCTGCGCGGAG 2250

RESULT 5
 AAT04785
 ID AAT04785 standard; DNA; 1145 BP.
 XX
 AC AAT04785;
 XX
 AC
 XX
 DT 17-JUN-1996 (first entry)
 XX
 DE Cellulase gene fragment 01-05.
 XX
 KW Cellulase; beta-glucosidase; cellulose; waste-disposal; ethanol;
 KM Escherichia coli; Saccharomyces cerevisiae; ds.
 OS Cellulomonas diazotica.
 XX
 FH
 FT
 FT CDS
 Key Location/Qualifiers
 331..1145
 /tag= a
 GB2289050-A.
 PN
 PD 08-NOV-1995.
 XX
 PF 05-MAY-1995; 95GB-0009237.
 XX
 PR 06-MAY-1994; 94GB-0009030.
 XX
 PA (UYHK-) UNIV HONG KONG.
 XX
 Chan WK, Wong WK;
 DR WPI; 1995-368626/48.
 PT New cellulase from Cellulomonas diazotica and related nucleic acid
 PT - used to degrade cellulosic waste, esp. to ethanol in conjunction
 PT with yeast glucanase(s)
 XX

PS Disclosure; Fig.14; 41pp; English.

XX The 0.75-kb pST1 and the 3.05-kb pST-Ndel fragments of *C. blazotea*
 CC ATCC 486 chromosomal DNA were sequenced using the dideoxy method. 8
 CC sections of the coding strand were identified, the first 5 of which
 CC were contiguous (given in AA04785) and included a putative start
 CC codon. The sequences of the other 3, non-contiguous, downstream
 CC fragments are given in AA07391-93. Cloning of the gene region allowed
 CC extracellular prodn. of recombinant cellobiose in *E. coli* and *S.*
 CC cerevisiae.

XX Sequence 1145 BP; 152 A; 444 C; 391 G; 158 T; 0 other;

XX Query Match 6.0%; Score 177.4; DB 16; Length 1145;
 XX Best Local Similarity 57.2%; Pred. No. 8.6e-20;
 XX Matches 362; Conservative 0; Mismatches 266; Indels 5; Gaps 2;

QY 447 GGAGTCCCTCTCTCCGCTTACAGATGGCCCAACGCGCTAAGAGGACCAAGTTCTTC 506
 DB 448 GGCTTCGGCAGAGTCCGCTGTCGACGCGCCGACGCGCTGCGGCTCAAGTTCTTC 507
 QY 507 AATGGCGTCCCTGGGCGCTGCTCCCTTGGCGACATGCTGCTGCTCAATCAACCA 566
 DB 508 GCGCGCGCGCAGCTGCTGCTTCCCAACGCGACCTGCTGCTGCTGCTGCTGCTGCTG 567
 QY 567 ACTGCTCGAAGAGCGAGTATGATGAGGGAAGGCGCATGCTTAAAGTGGCAT 626
 DB 568 GAGTCCGACGAGAGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 627
 QY 627 GTGATCTCGCGCGCGCATATACATGCAACGCTCCCTCGGTTGGAGCGCTTCGAG 686
 DB 628 GTGCTGCTCGCGCGCGCATATACATGCAACGCTCGCTGCTGCTGCTGCTGCTGCTG 687
 QY 687 TCGATTTGGAGAGATCGCTTCCTGCGCGCTTGGAGAGCTGCGGCTCTCATCCGCGCAT 746
 DB 688 GCTTACTCGAGAGACCGCTGCTACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 747
 QY 747 CAGAGCACTGAGT-GCAGGCTACGATACACTTTTGGCAATGATCAAGAGAGACAG 805
 DB 748 CAGGACCTCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 807
 QY 806 GGGCAATGATGCTGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 865
 DB 808 GCGCAACACCATGAACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 867
 QY 866 GTTCAGATTTGCTGCGAGACTCCAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 925
 DB 868 GTTCAGATTTGCTGCGAGACTCCAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 927
 QY 926 CATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 985
 DB 928 CAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 987
 QY 986 GGGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1045
 DB 988 GGGCTACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1047
 QY 1046 TGTGGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1078
 DB 1048 CG-----CGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1076

RESULT 6
 AA87294
 ID AA87294 standard; DNA; 2430 BP.

XX AA87294;
 XX AC
 XX 05-JUN-2000 (first entry)
 XX DT
 XX S. venezuelae macrolide beta-glycosidase gene desR, SEQ ID NO:23.
 DE
 XX Desosamine biosynthesis; macrolide; polyketide; pikromycin; pikromycin;

KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolemia; crop protection agent; ds.
 OS Streptomyces venezuelae ATCC15439.

XX Key location/Qualifiers
 XX CDS 1..2340
 XX /tag= a
 XX /product= "DesR"

PN WO20000620-A2.

PD 06-JAN-2000.

XX 25-JUN-1999; 99WO-US14398.

XX 26-JUN-1998; 98US-0105537.

XX (MIND) UNIV MINNESOTA.

XX Sherman DH, Liu H, Xue Y, Zhao L;

XX WPI; 2000-160679/14.

XX P-SDB; AA77189.

PT Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 PT synthesis of methymycin and pikromycin

PS Claim 3; Page 369-370; 438pp; English.

XX The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the *eryC* gene cluster of *Saccharopolyspora erythraea* or
 CC Streptomyces antibioticus. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
 CC narbomycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthesis may be useful to prepare novel antibiotics and
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biomedical applications, to engineer PHA monomer syntheses or to prepare
 CC biologically active agents, such as chemotherapeutics,
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants.
 CC Sequences AA87286-287294 represent desosamine biosynthetic genes from
 CC Streptomyces venezuelae ATCC 15439, which encode proteins
 CC AA77181-77189.

XX Sequence 2430 BP; 407 A; 878 C; 834 G; 311 T; 0 other;

XX Query Match 5.9%; Score 175.2; DB 21; Length 2430;
 XX Best Local Similarity 47.6%; Pred. No. 2e-19;
 XX Matches 1016; Conservative 0; Mismatches 968; Indels 150; Gaps 11;

QY 544 CGCTGGTTCACATCAACCAACTGCTGGAAGAGCGATGAATGATGGCAAG 603
 DB 353 CCCTGGCGAGACCTTCGACGACCATGCGCAGAGTACGCGAAGTCAAGGCGCGC 412
 QY 604 AGGCATCGCTTAAGATGGGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 663

Db 413 ACGTCGCGCTCAACAGACATGTCCTGGGCGCGATGATGACATCCGGGTGC 472
 Oy 664 CTCTGCTGAGAGCTGCTTCAGATGATTTGGTGGAGATCCGTTCCCTGGCGGCTTGGAG 723
 Db 473 CGCAGCGGGGCGGACATCGAGACCTTTCAGCGAGACCGCCCTGCTCTCCGCGACCG 532
 Oy 724 CTGCGGCTCTCACTCCGCGCATTCAGAGCACTGAGTGCAGGGTACGATCAACACTTTT 783
 Db 533 CGGTGGCCGATCAAGGCGATCCAGGGGTGGGCTGATGACACGCGGACACGACTTGC 592
 Oy 784 TGTGATGATGATGAGAGGAGGCGATGATGAGAGTGCAGAGCTGCTACGAGCGGGCTC 843
 Db 593 CGGCGCAACACGAGAGGAGACACGCTTCTCCGTAAGCGCCATGTCGACACACACCG 652
 Oy 844 TCCGTAATCTACGACATCCGCTCCGATTCAGATTGCTGTGAGAGTCCGACCGGGTGCCT 903
 Db 653 TCCGAGATGATGATTCCTCCGCGCTTCAG---GGTCTCCAGCGCGCGGCGGCTCTCT 709
 Oy 904 TCATGAGCGGCTACATGAGGATTCATATGCGCTGTGACGAGAACCTTAATATTTG 963
 Db 710 TCATGCTGCTACACGCGCTCAACGGAAGCGCTCCGCGCACGACGACGACTCTCA 769
 Oy 964 ATGGATGCTTCAAGAGAAATGGGTTGGGATGGCTTATCATGAGCGAGTGTAGGCA 1023
 Db 770 ACAAGTGTGCGCAGCAGAGTGGGCTTCCAGGGCTGGGATGATTCGACTGCTGCGCA 829
 Oy 1024 CATACGATACACAGAGCGCTGTGTGAGGCGCTCACTCGAGATGCCGAGCTCCAC 1083
 Db 830 C---CGGCGGACGAGCGCATCACAGGCGCTCGACAGAGATG--- 873
 Oy 1084 GCTTCGAGAGAGAACTCAAGTTCACGCTCCAGAGGAAAGCCCTTATCCAGTCA 1143
 Db 874 -----GGGTCGAGCTCCCGGCGAGCTCCGAGAGGCGAGCCCTCCGCGCGCA 925
 Oy 1144 TTGACGAGAGGCTAGGGAATTCCTCACTTCGTAAGAAATGTGCTGCTCGGAGTGA 1203
 Db 926 AGTTCTTGGCGA---GGCGCTGAGACGCGCGCTCTGAAAGCGAGCGTCCGAGGGGG 982
 Oy 1204 CGGAGAGCGCGCGAGAGAGCTGCTGTAAGACACACCGCGGAAAGCGAGCTCTCCGGA 1263
 Db 983 CCGTGAAGCGGTGGGAGAGCGGATGTCGCGCAGATGGAAGTTCGCTGCTCTCG 1042
 Oy 1264 AGTTGGCAAGAGGCGATCTGCTGTAAGAGAGAGAACTTCGCGCTTGAGCA 1323
 Db 1043 CCAGTCCGCGCGCGCGCGAGCGGAGACAGCGGCGGTGCGCGAGTCCGCAAGG 1102
 Oy 1324 AGAAGAAAGAGAGCTGATGTCGCGCCCAAGCGCAAGCGGCTCATACAGCGAG 1383
 Db 1103 TCGCCGAGAGCGCGCGGTCTCTGCGCAGAGAGGCGCGCTGCGCGGTG 1162
 Oy 1384 GCTCTGCGCGACTAGGCGCTCTACTAGCGAGTCACTCCCTTGAGCGCTTCAGAGAGC 1443
 Db 1163 AGCGCGGAGAGAGATCGGGGCTATCGCGCGAGCGCGCTGAC---CCAGAGTCAAGCGG 1221
 Oy 1444 TCGAGAGCGCGCGCATGTAACCGTCCGCGCTTACACACCGCTTCCTCCATTTAGAGG 1503
 Db 1222 CTGGGAGAGCGCGCGTCTCCGAGCTGGGCGGCGCGCACTGACACCATAGAGCC 1281
 Oy 1504 AGCAGTGCCTCAAGCGCGCGCGCTCGGGGATGCGGTGAGAGGCTTCACAGAGCC 1363
 Db 1282 CCGCGGGGTGCGGGTGCAGGCTGACGTAGAGAGAGGGGTGAGAGCTTCGAGAGCCAG 1341
 Oy 1564 CTGGTACCCCTTAACCGAGACATTTGAGAGCTCTTCTTACACAGAGAGATGAGAC 1423
 Db 1342 ATCCGCGGGGAGACCTGAGC-----CGGCGCTTCAAC 1375
 Oy 1624 TGTGAGTACTACTACACCCCAAGCGGAGACAGCTGTAGAGCGCGAGATGAGAGCACT 1683
 Db 1376 AGGCGCCAGAGCTCAGAGCGGAGCAAGGCGGGGCTGTAGAGAGCGCACTGACGCTG 1435
 Oy 1684 ACACCGCGAGAGAGTCACTAGAGCTGCGGCTGCTGCTGTGAGAGAGCGCAAGG 1743
 Db 1436 CCGCGAGGCGGAGTACCGCATCCGCGTCCGTCACCGAGGTGTTACGCGACGCTGACG- 1494

Oy 1744 GGTACTGAGACGACGCTCTGCTGACAAAGCCACAGAGATCCCGGATGCT 1803
 Db 1495 -----CTCGGAGGCACA 1507
 Oy 1804 TCTTGGCTCCGCGCACCGCGAGAGAGCGCGCATCAATCTGTCAGAGGCGCAACGT 1863
 Db 1508 CCATGAGGCGGCTGAGTCTGCGCAAGAGTGAAGCACCGCGCTCTCAAGCTGACCAAG 1567
 Oy 1864 ACAAGTTCAGATGAGATGAGTGGCTCCGACACCACTTACACCTTCAAGGGGAGACCATG 1923
 Db 1568 GCAGGACAAAGCT-----CAGATCTGCGGCTTGCATTA 1603
 Oy 1924 TCCCGGCGAGCGCTCCCTCCGCTCGCGGCTGCAAGTCAATTAAGACAGCGCGAA 1983
 Db 1604 GTGCACCGCGCTCTCTGAGAGTGGGCTGGGTGAGACGCGCGCGCGCGCGAGCA 1663
 Oy 1984 TCGAAGATCGCGCTCCGCGCGCAAGAGAGCAGCAGGATCATCATCTGCGGGCGCTTA 2043
 Db 1664 TCGGGAAGGCGCTGAGTCCGCGGAGAGCGCGTACGCGGCTGCTTGC----- 1715
 Oy 2044 ACGCCGACTGGAGACCGAGGCGCGACCGCGAGCATGAAGTCCCGCGCTGTG 2103
 Db 1716 -CTAGACAGACGCGACGAGGCGGTGACCGTCCGAACCTGTGCGTCCGCTGAGCAG 1774
 Oy 2104 ACCAGCTCATTTGCGGAGGCGCGCGCAACCGCAACACCGCTGCTCATGAGAGG 2163
 Db 1775 ACAAGCTGATCTGGCTGTGCGCGAGCGCCACACCGCAAGATGCTGTCTCAACACCG 1834
 Oy 2164 GCACCGCGGAGAGATCCCTGCTGAGCGCACGCGCGCGCTCATACAGCGCTGTAGC 2223
 Db 1835 GTTCGTGGGTGTGATGCGGTGCTCCAGAGCCCGCGGTCTTGAGCATATGTATAC 1894
 Oy 2224 GCGGCAAGGAGAGGCGCACTCATTTGCGGAGTGTCTTTGCGCACTCAACCTCTCG 2283
 Db 1895 CGGCGAGGCGGCGCGCGAGCGCACCGCGCGCTCTCTACGATGATCAACCGAGCG 1954
 Oy 2284 GCAAGTGTCTTCCAGCTTCC-----AAGCGCTGAGAGCAACCGCGTTTC 2334
 Db 1955 GCAAGCTCACGAGACCTTCCGCGCGCGCGAGAACACAGCGCGGTCCCGGACCGA 2014
 Oy 2335 TCACTTCCGACCGAGCGCGCGCGAGCTGTACGCGGAGAGCTTACGCTGCGGTA 2394
 Db 2015 CAAGCTACCGGCGGTGCAACAGACAGACGTACCGGAGGATCCAGCTCGGTAAC 2074
 Oy 2395 GGTACTACGAGTTTGGCGCAAGGAGCGTCAATTTCCCTTGGCGAGCGCTGTCTACA 2454
 Db 2075 GCTGGTTCAGCAAGAGAACTCAAGCCGCTGTCTCCGCTGCGAGCGCTGTCTACA 2134
 Oy 2455 CCAGTTTCCCTTTCATCTCTCGTCTGCTCAAGAGAC---GGCAAGCTGAGCGT 2511
 Db 2135 CCTGCTTACGAGAGCGCGCGCGCGCGTGTGCTGAGTCCAGGAGTGTGAGAGTCA 2194
 Oy 2512 CCCTTCCGTAAGAAAGCGGCTCCGCGCGCGCGAGAGTGTCCGAGCTTACGTCA 2571
 Db 2195 CGGTACGCTCCGCAACAGCGGGAAGCGCGCGCGAGAGTGTCTGAGAGCTTAC 2254
 Oy 2572 AGCCCTCCAGCGCGCGCAAGTTAAGCGCGCGCTCAAGAGAGCTCAAGGCTTGGCAAG 2631
 Db 2255 GTGCCAGCGCGAGAGTGAAGGCTCCGAGGAGGAAGAGTGTGTGAGTACAGAGG 2314
 Oy 2632 TCGAAGTCAAGCGCGCGAGAGAGAGGCTGAC 2665
 Db 2315 TCTGCTGCGCGCGCGAGCGAGCGAGAGGTAAC 2348
 RESULT 7
 AAD39052
 ID AAD39052 standard; DNA: 2430 BP.
 AC AAD39052;
 AC
 AC
 DT 23-SEP-2002 (first entry)

XX Streptomyces venezuelae Desr gene.
 DE glycosylated polyketide; modified recombinant bacterial host cell;
 KW mRHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;
 KW tetracycline; polyene; polyether; ansamycin; isochromanone; sugar;
 KW desosamine; Desr; glucosidase; enzyme; gene; ds.
 XX Streptomyces venezuelae.
 OS
 XX Key Location/Qualifiers
 FH 1..2430
 FT CDS /tag= a
 FT /product= "Streptomyces venezuelae Desr protein"
 FT /transl_except= (pos:1..3, aa:Met)
 FT /note= "CDS does not include start codon"
 FT /partial
 XX WO200229035-A2.
 XX 11-APR-2002.
 XX 05-OCT-2001; 2001WO-US31255.
 XX 05-OCT-2000; 2000US-238185P.
 XX (MINU) UNIV MINNESOTA.
 XX (LIUH/) LIT H.
 XX (SHER/) SHERMAN D H.
 XX (ZHAO/) ZHAO L.
 XX Liu H, Sherman DH, Zhao L;
 XX WPI: 2002-405171/43.
 DR P-PDB: AAE24237.
 XX Modified recombinant bacterial host cells in which the expression and
 PT activity of nucleic acids encoding sugar biosynthetic enzymes has been
 PT altered, useful for producing metabolites with altered sugar structures
 PT
 XX Disclosure; Page 170-171; 174pp; English.
 PS The invention provides a method to alter the sugar structure diversity
 XX for a particular metabolite via the recruitment and collaborative action
 CC of sugar genes from a variety of sugar biosynthetic pathways to yield a
 CC metabolite comprising a non-natural sugar, e.g., a novel glycosylated
 CC polyketide. The invention also relates to a modified recombinant
 CC bacterial host cell (mRHC) in which the expression and activity of
 CC nucleic acids encoding sugar biosynthetic enzymes has been altered.
 CC The mRHCs may be cultured to produce the modified sugar products,
 CC e.g. a macrolide, anthracycline, angucycline, avermectin, milbemycin,
 CC tetracycline, polyene, polyether, ansamycin or isochromanone.
 CC The present sequence is Streptomyces venezuelae sugar (desosamine)
 CC biosynthetic gene cluster Desr (glucosidase) gene.
 XX
 SO Sequence 2430 BP; 407 A; 878 C; 834 G; 311 T; 0 other;

Query Match 5.98; Score 175.2; DB 24; Length 2430;
 Best Local Similarity 47.68; Pred. No. 2e-19;
 Matches 1016; Conservative 0; Mismatches 968; Indels 150; Gaps 11;

OY 724 CTGCGGCTCTATCCGGCGCATTCAGAGCACTGGAGTGCAGGCTACGATCAACACTTTT 783
 DB 533 CGGTGCGCCAGATCAAGGGCATCCAGGGGTGGGTCTGATGACACCGCCAAACACTTGG 592
 OY 784 TGTGCAATGATCAGAGAGCAGGCGCATGTGTGACAGACATCTGACAGGCGGCTC 843
 DB 593 CGGCCAACAACAGAGAGAGCAACACGCTTCCGGAAGCCCAATGTGACAGCAACG 652
 OY 844 TCCGTGAATCTACGCACTCCGTTCCAGATTTGTGTGCGAGACTCCACGCGGTGCT 903
 DB 653 TCCGCGAGATCGAGTTCCCGGCGTTGAG---GCTCTCCAAAGCGCGGCGCTCTC 709
 OY 904 TCATGACGGCGCATCATGATCATGATGCTGTGTCAGAGAGAACCCCAATATCTTG 963
 DB 710 TCATGTGTCTTCAACAGGCTCAACGGAAGCCGCTCTGCGGCAAGAGAGCTCTCA 769
 OY 964 ATGGGATGCTTGAAGAGAAATGGGTGGATGCGCTTAATATATAGAGAGCTGTAGGCA 1023
 DB 770 ACAAGGTGCTGCGACAGAGATGGGGCTTCCAGGGCTGAGATGCTCGACTGGCTGCCA 829
 OY 1024 CATACAGTACCAACAGAGCGCTTGTGCGAGGCTGACCTGAGATGCCGACTCTCAC 1083
 DB 830 C---CCGGGCGACCGAGCCATCACCAAGGCGCTCGACAGAGATG----- 873
 OY 1084 GCTTCCGAGAGAAACACTCAAGTTCAACGTTCTCCACGAAAGCCCTTTATCCAGTCA 1143
 DB 874 -----GGGTGCAAGTCCCCGGGAGCTCCGAAAGGGGAGCGCTCGCGCGGCGCA 925
 OY 1144 TTGACCAAGAGGCTAGGGAAGTTCTTCAGTTGTCAGAAAGTGTCTCTCCGAGTGA 1203
 DB 926 AGTTCTTTCGGCGA---GGGCTGAAGAGCGGCTCTCTGAAGCGGCTCCGAGGGGG 982
 OY 1204 CGGAGAAGCGCCCGGAGAGAGCTGTCAACACACCCCGAAAGCGGAGCTCTCTCCGGA 1263
 DB 983 CCGTGACGCGGTGGGGAGGCGGATGTGCGCCAGATGAGAAATGTGCTGTCTCG 1042
 OY 1264 AGTTGGCAACAGGCGCTGTCTCTGAAGAAGAGAGAAACAGCTGTGCTCTGACCA 1323
 DB 1043 CCACTCGGGCGCGCGCGCGAGCGGACGAGAGGCGGGTCCAGGCGGTGCTCCCAAG 1102
 OY 1324 AGAAGAAAGAGAGCGTGTATGTCGCGCCCAAGCGCAACATACACGCGCGAG 1383
 DB 1103 TCGCCGAGAAAGCGCGGCTCTCTGCGACAGAGGCGCCAGCGCTGCGCGGTG 1162
 OY 1384 GCTTGTCCGACCTCAGGCGCTTACTACGAGTCACTCTTTAGAGGCGCTCAGCAAGCAG 1443
 DB 1163 ACGCGGCAAGAGCATGCGGTATCGGCGCGAGCGGCGCTGAC--CCAAAGTACCGGC 1221
 OY 1444 TCGAGAGCGCGCATCTGTACACGCTGGGCGCTACACCAACCGTTCTCTCCATCTAGGCG 1503
 DB 1222 CTGGGCAAGCGCGCTGTCTGCGAGCTGCGGCGCGGCGGCGCACTGACACATCAAGGCC 1281
 OY 1504 AGCAGTGCCTACCGCGAGCGGCGCTCGGGCATGTGCGTGGAGGCTTTCACAGAGCC 1563
 DB 1282 CGCGGGGTGCGGGTGGACGAGTACGTACGAGACGGGTGAGAGCACTTGGGAGCGAG 1341
 OY 1564 CTGTGATCCCTTAACCGCGACATGTAGAGGCTCTTCAACCAAGAGAGGAGATGACCC 1623
 DB 1342 ATCCCGGGGGGAACCTTAGC-----CGGCGGTCAACC 1375
 OY 1624 TGTGTGACTTACTACACCCCAAGGCGGAGAGACGTGTACCGCGACATGAGAGGACGT 1683
 DB 1376 AGGGCCACACACTGAGAGCGGAGCAAGGCGGGGCGGTGTACGAGCGGACCTACCGTGC 1435
 OY 1684 ACACCGCGGAGAGAGTGTACCTACAGGCTGCGCTGTGCGGCGAGCGCAAGG 1743
 DB 1436 CGCGCGAGCGGAGTACCGCATCGCGGTCTGTGACACCGGTGTGTAGCGACGAGTGCAG- 1494
 OY 1744 GTTACGTAGAGACAGTCTGTGTGCAACAGCCCAAGAGAGAGTCCCGGAGATGCT 1803
 DB 1495 -----CTGGGAGGCCACA 1507

QY 1804 TCTTGGGCTCCGCCACCCGCGAGAGACGGCCGATCATCTGTCAGAGGCAACAGT 1863
 DB 1508 CCATGAGAGCCGGTCAAGGTCTACAGGAGTGAGCAGCCCGCTCTCAAGCTGACAGG 1567
 QY 1864 ACAATGTAAGATCGATGTTGGCTCCGACCCACTTACACCTCTAAGGGGACACCAATCG 1923
 DB 1568 GCACGCACACAGT-----CAGATCTCGGGCTTCGCATGA 1603
 QY 1924 TCCCGGCGCAGGCTCCCTCCGCTCGAGCGCTGCAAGTCAATTGACAGCAGCCGAAA 1983
 DB 1604 GTGCACCCCGCTCTCCCTGAGCTGGGGGTGGAGACCGCGGGGCGCGACGACCA 1663
 QY 1984 TCGAAAGTCCGTGGCCCTCGCCAGAGACACACAGTCAATCTGCGGGGCTTA 2043
 DB 1664 TCGGAAGGCCGTGGAAGTGGCGGAAAGGCCGTACGGGCGTCTTCG----- 1715
 QY 2044 AGCGGACTGGAGACCGAGAGCGCGCGCGCGGAGCATGAAGCTCCCGGCTGCTGG 2103
 DB 1716 -CTAGGACAGGAGCGGACGAGGCGCTGACCGTCCGAACTGTGCTGGCGGGTACGCAAG 1774
 QY 2104 ACAGCTCATTTGCCGAGCTGAGCGCGCGCGCAACCCAAACACCGCTGCTATGACAGCG 2163
 DB 1775 ACAAGCTGATCTCGGCTGTGCGGAGCGCAACCGGAACAGATCGTGTCTCAACACCG 1834
 QY 2164 GCACCCCGGAGAGATGCGCTGAGCGCACCGCCCGCTCATTCAGCGCTGGTAGG 2223
 DB 1835 GTTGTGGTGTGCTGATCGCTGGCTGTCCAAAGACCGCGGCTCTGGAGCATGTGTACC 1894
 QY 2224 GCGGCAACGAGAGCGGCACTTCATTGCGGAGCTGCTTTGGGAGCTAACCCCTCGG 2283
 DB 1895 CGGGCCAGGGGGGCGGAGCGCCGCGCTGTCTACGCTGACGTCACGTCAGCG 1954
 QY 2284 GCAAGCTGCTCTCAGCTTCC-----AAGCGCTGAGAGCAACCCCGCTTTC 2334
 DB 1955 GCAAGCTCAGCAGACCTTCCGCGCGCGAGAGACCGAGCAGCGTCCGCGGCGACCGCA 2014
 QY 2335 TCAACTTCCGACGAGCGCGGCGCGAGCTGACGAGGAGGAGTCAAGCTCGGGTACA 2394
 DB 2015 CAAGCTACCGCGGCGTGCAGCAACACAGACAGTACCGGAGGCACTCAGCTCGGGTACC 2074
 QY 2395 GGTACTACGAGTTGCCGCAAGAGACGTCAATTCCCTTTGGCCAGCGCTGTCTACA 2454
 DB 2075 GGTGGTTCGCAAGAGAACTCAAGCCGCTTCCGCTGGGCGAGCGCTGTGTACA 2114
 QY 2455 CCACTTTCCTTTTCACTCTCTCGGTGTCTCAAGAGC---GGCAAGCTGAGCTGT 2511
 DB 2135 CCGTGTTCACGAGAGAGCGCCGACCGTGTGCTGATCTCAAGGCTGTGTAGGTCA 2194
 QY 2512 CCTCTCGTGAAGAACACCGGCTCGGTCCGCGGCGACAGCTGGCCAGCTTACGTCA 2571
 DB 2195 CGGTACGCTCGCAACAGCGGAGCGCGCGGCGAGAGTGTCTCAGCGCTTACTCG 2254
 QY 2572 AGCCCTCCAAAGCGGCGCAAGTAAACCGCCCGCTCAAGAGCTCAAGGCTTCCGAAGG 2631
 DB 2255 GTGCCAGCGCGGAGCTGAGCGGCTCGCAGGCGAAGAAAGCTGTGGGTACAGGAAG 2314
 QY 2632 TCGAACTGCAAGCCGCGGAGAGAGAGGCGGTGAC 2665
 DB 2315 TCTCGTCCGCGCGGCGAGGCGAAGAGCGGTGAC 2348
 RESULT 8
 AA87284 standard; DNA; 12441 BP.
 XX AA87284;
 AC
 XX
 DT 05-JUN-2000 (first entry)
 XX
 DE S. venezuelae desosamine biosynthetic gene cluster pikB, SEQ ID NO:3.
 KW Desosamine biosynthesis; macrolide; polyketide; pikromycin; pikromycin;
 KW neomethylmycin; narbomycin; polynhydroxyalkanoate monomer synthase;

KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolemia; crop protection agent; ds.
 OS Streptomyces venezuelae ATCC15439.
 PN WO200000620-A2.
 PD 06-JAN-2000.
 XX 25-JUN-1999; 99WO-US14398.
 XX 26-JUN-1998; 98US-010537.
 XX (MNU) UNIV MINNESOTA.
 PA Sherman DH, Liu H, Xue Y, Zhao L;
 PI WPI: 2000-160679/14.
 DR P-PSDB: AAY77179.
 XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
 PT synthesis of methylmycin and pikromycin
 PS Claim 2; Page 281-287; 438pp; English.
 CC The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the eryc gene cluster of *Saccharopolyspora erythraea* or
 CC streptomycetes antibiotics. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC pikromycin, neomethylmycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the production
 CC of biologically active macrolides. The macrolide biosynthetic proteins
 CC are useful for synthesis of methylmycin, pikromycin, neomethylmycin and
 CC narbomycin. The alternative termination of polyketide synthesis may be
 CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
 CC monomers. The compounds produced by the recombinant host cells are useful
 CC as biopolymers, e.g., in packaging or biomedical applications, to
 CC engineer PHA monomer syntheses or to prepare biologically active agents,
 CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
 CC chronic obstructive pulmonary disease as well as other diseases involving
 CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
 CC antibiotics which are active against a variety of organisms, e.g.,
 CC bacteria, including multi-drug resistant pneumococci and other
 CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
 CC protection agents (e.g., fungicides or insecticides) via expression of
 CC polyketides in plants. The present sequence represents the desosamine
 CC biosynthetic gene cluster from *Streptomyces venezuelae* ATCC 15439.
 XX
 SO Sequence 12441 BP; 1704 A; 4294 C; 4686 G; 1757 T; 0 other:
 Query Match 5.9%; Score 175.2; DB 21; Length 12441;
 Best Local Similarity 47.6%; Pred. No. 2.2e-19;
 Matches 1016; Conservative 0; Mismatches 968; Indels 150; Gaps 11;
 QY 544 GCGTGGTTCCACATTCACAACTCTGCTCGAAGAGGACGTAAGATGATGGCAAG 603
 DB 3977 CCTTGCCAGACACCTTGACACACATGCGCCGACACTACGCAAGGTATGGCGCGG 4036
 QY 604 AGGCATCGCTTAAAGATGCGCATGTGATCTCGGCGCGACTATACATGCAACAGCTCCG 663
 DB 4037 ACGGTGCGGCGCTCAACAGGACATGTCTCTGGCCCGGATGATGAACAACTCGGGTCC 4096
 QY 664 CTCTCGGTGACGTGCTTGAAGTCAATGTTGTTGTTCTGCGGGCTTGGAG 723
 DB 4097 CGCAGGCGCGCGCAACTTACAGCACTTACAGCAGACCCCTGTCTCTCGGCGACCG 4156
 QY 724 CTGCGGCTTCATCCGCGGATTCAGAGCACTGAGTACAGCTTACATCAAGCACTTTT 783

OY	1684	ACAAGTTCAAGATCGAGTGTGGCTCCGACACCACACTTCACACCTTCGAAGGCGGCACCATCG	1923
Db	51192	GCACGCCAAGAAGCT-----CAGATCTCGGGCTTCGCATTA	5227
OY	1924	TCCCGGCCCCAGGCTCCCTCCGCGTGCCGCGCTGCAMAGTCAATTATTCAGCACGACGAGCCGAAA	1983
Db	5228	GTCGCACACCCCGCTCTCCCTGGAGCTGGCTGGGTGGTAGCCGGCGGGCCGAGCGCAGCA	5287
OY	1984	TCGAAAAAGTCCGTGCGCCCTCGCCAAGGACGACGACAGTCATCATCTGTGGCGGGCCCTTA	2043
Db	5288	TCGGGAAGGCGCGTGGAGTGGGGCGGAAAGGCCCGTACGGCGGTCTGTTTCGC-----	5339
OY	2044	ACGCGCACTGGGAGACGAGGGCGCCGACCGCGAGCATTAAGCTCCCGGCGTCTGG	2103
Db	5340	-CTACGACGACGGGACACCGAGGGCGCTGCAGCCTTCGMAACTGTGCTCTCCGGGTAGCGAAG	5398
OY	2104	ACCAAGCTCATTTGCGCAGGTGGCCGCGCGAACCACCAACCCGCTGTCGTATCGAACGG	2163
Db	5399	ACAAGCTGAATCTGGCTGTGCGGCGAGCGCAACCCGCAACAGCATGTTGTGTTCMAACCG	5458
OY	2164	GCACCCCGGAGGAGTGGCCCTGGCTCGAAGCGACGCGCGCGGTATCGACGAGCCGTGGTAG	2223
Db	5459	GTTCGTGGGTGATGATCCGTGGCTGTCCAMAAGCCGCGCGCTCTGGAATGTGGTACC	5518
OY	2224	GGCGGACGAGAGAGGGGCAACTCCATTGCGCAGTGTGTTTTGGGACTATMACCCCTCGG	2283
Db	5519	CGGGCGCAGGGGGGGCGCGAGGCGACCGCGCGCTCTGTACGTGATCGTCAACCCGAGCG	5578
OY	2284	GCAAGCTGTCTCTACGCTTCCC-----AAGCGCTCAGAGAACACCCCGCGCTTC	2334
Db	5579	GCAAGCTCAGCAGAGACTTCTCCGGCGCGGACGAAACACACGCGGTGCGGGGAGCCCA	5638
OY	2335	TCAACTTCCGACCGAGGCGCGGCGACGACTGTACGGCGAGAGAGTCTACGTGCGGTACA	2394
Db	5639	CAAGCTACCGGGGCGTGCACAAACGACGACGTAACCGGAGAGGATACGATGGGGTACC	5698
OY	2395	GGTACTACGAGTTTGGCCGACAGAGAGCTCAATTTCCTTTGGCCACGCGCTGTCTTACA	2454
Db	5699	GCTGTTTCGACAAAGAGAAAGCTCAAGCCGCTGTTCCCGTTGGGGCAGCGCTGTCTGATCA	5758
OY	2455	CCACTTTTGGCTTTTCCATCTCTCCGCTGTCTCAANAAGAC--GGCAAGTGGCGGTG	2511
Db	5759	CTCTGTTTCACGACAGAGGCGCCGACCGTGTGCTGATCTCCACGAGGTGGTGTGAAGTCA	5818
OY	2512	CCCTCTCCGTGAAGAACACCGCGCTCCGTCGCCGCGCACAGGTGGCCACGCTCAAGTCA	2571
Db	5819	CGGTACGAGTCCGCAACAGGGGGAACGGCGCGGCGCAGAGAGTGTGTGAGGCGATCTCG	5878
OY	2572	AGCCCCCTCCAAAGCGGCCAAGATTAAACCGCCCGTCAAGSAGCTCAAGGCTTGGCAAAGG	2631
Db	5879	GTCGCACCGCGAAGTGAACGCGCTCCGACGACAMAGMAAAGCTGTGGGCTTACACGAAGG	5938
OY	2632	TCGAAACTGACAGCCCGGCGAGACGAAGCGGTGAC	2665
Db	5939	TCTGCTGCGCGGGGCGAGGCGAAGACGCTGAC	5972
<hr/>			
RESULT 9			
AAD39043			
ID	AAD39043	standard; DNM; 13613 BP.	
XX	AAD39043:		
AC			
XX			
DT	23-SEP-2002	(first entry)	
XX			
DE	Streptomyces venezuelae desosamine gene cluster.		
KX	Glycosylated polyketide; modified recombinant bacterial host cell;		
KW	mRHC; macrolide; anthracycline; angucycline; avermectin; milbemycin;		
KW	tetracycline; polyene; polyether; ansamycin; Isochromanepunone; sugar		
XX	desosamine; des; gene; ds.		

CC	module 1-4 or 1-6, including an oleandrolide PKS operably linked to a
CC	Promoter. Also discussed are recombinant oleandrolide PKS in which the
CC	module 1 KS domain is inactivated by deletion or other mutation. In
CC	particular, the inactivation is mediated by a change in the KS domain
CC	that renders it incapable of binding substrate (the KSI-o mutation),
CC	rendered by mutation in the codon for the active site cysteine. The
CC	oleandrolide PKS is useful for synthesizing polyketides, which are useful
CC	as antibiotics and molluscs. Heterologous expression of oleandrolide PKS
CC	in host cells such as Streptomyces coelicolor and S. lividans is also
CC	made possible. Unmodified oleandrolide compounds can be provided to
CC	cultures of Saccharopolyspora erythraea and converted to the
CC	corresponding derivatives of erythronmycins A-D.
XX	
SQ	Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;
Query Match	4.7%; Score 136.6; DB 21; Length 50937;
Best Local Similarity	54.0%; Pred. No. 1.8e-13;
Matches 354:	Conservative 0; Mismatches 289; Indels 12; Gaps 3
QY	2048 CGACTGAGGAGACAGGGGCCCGCAGCGCGAGACATGAAGCTCCCCGGCGTGCAGCA 2107
DB	3417 GCAGCAGCGGTCTGGAGAGACCGGATCCGCATGTCCTGTCTCCCGCGCCAGGACGA 3358
QY	2108 GCTCATTTCCGACGTGGCCCGCCGCGAACCCAAACCCTGTGTCATGCAGACGGCAC 2167
DB	3357 CCTGTATGACGCCCTGTGGCCCGCGGTCAACACGACGCCGTHGGTGCTCAACACGGGCTC 3298
QY	2168 CCCGAGGAGATGTGCTGCTGACGCCCAACGCCGCCCTCATCCAGGCTGTGTAACGGCG 2227
DB	3297 CTCCTTACCATGCGCTGCTGAGAAAGACCGCCGCTGTCTACCATGTGTGTAACCGGG 3238
QY	2228 CAACGAGACGGGCAACTCTATTCCCGACGCTCTCTTTGGCATACAACCCCTCGGGCAA 2287
DB	3237 CCAGCGCGCGCGCGAGGCGACACACGCCCTGTGTGTGGCGACGCCACCCAGAGGCGG 3178
QY	2288 GGTGTCCTTAGCTTCCC-----CAAGGCGCTGCAGACAACCCCGCTTTCTCAATT 2341
DB	3177 GCTGACCCGACACTCTTCCCGGCGACGAGGGGAGACGCCCTTCCCGGCGACGCCAGCGG 3118
QY	2342 CCGGACGAGGCGCGGCGGCGACGCTG---TACGGGAGGAGACGCTAAGTGGSGTACAGGTA 2398
DB	3117 CTACCGCGGAGGTGAGTGAACCACTGGACTACTCTCGAGGGGCAATTCTATCTCGGCTPJC 3058
QY	2399 CTACGAGTTTGGCCGACAGAAGACGTCAATTTCCTCCCTTGGCCACGCGCTGTCTACACCAC 2458
DB	3057 GTACGAACACAGACGAGGCGGTCCAGCCGCTGTCTTCCTTGGCCACGAGGTTGTGATCAACCTC 2998
QY	2459 TTTTGGCTTTTCCAATCTCTCCGTGTCTACAAA---GAGGCGCAAGCTGACGCTGTCCCT 2515
DB	2997 CTTGACTATACCGGACCTTAAGGTATACGACGACGCGGCGGGGTCTGACTGTCTGCTT 2938
QY	2516 CTCGCTGAMAGAACACGCGGCTCCGTCCCGGCGGACAGAGTGGCCAGCTGTACGTCAAGGC 2575
DB	2937 CAGCTCTCGGACACACCGGTACGCGGACCGGCAAGAGAGTCCCCGAGTCTACTGTGGGCGC 2878
QY	2576 CTCGCAAGCGGCCAAGATTAAACCGCCCCGTCAAGGACTCAAGGGCTTGCAGAAAGTCTGA 2635
DB	2877 GTCCCGGACAGTGTGCGGTGGCGCAGAGCCAAACGCTGCCCTGTGGCGGCTTACGGGAAGTCTGA 2818
QY	2636 ACTGACAGCGCGGAGAGAGAGAGGGGTGTACATGTAGAGAGCAGAGAAATACTCT 2690
DB	2817 GCTGCGCGCGGCGAGGACGAGAGGCTGACCTGTGCACAGTGCAGCGCAGGGCCCTC 2763
RESULT 13	
AC	AAT93682
ID	AAT93682 standard; DNA; 2166 BP.
XX	AAT93682;
XX	12-MAR-1998 (first entry)
DE	Thermoplasma maritima MSRA glycosylase accession DNR

XX glycosidase; thermostable; textile; food processing; pharmaceutical;
 KM detergent; baking; industry; Thermococcus; Staphylothermus;
 KM Pyrococcus; glucose; soluble oligosaccharide; ss.
 XX Thermotoga maritima.
 OS
 FH Key Location/Qualifiers
 FT CDS 1..2166
 FT /tag- a
 FT /product- glycosidase
 XX MO9725417-A1.
 XX 17-JUL-1997.
 XX PD 10-JAN-1997; 97MO-US00092.
 XX PR 13-SEP-1996; 96US-0712612.
 XX PR 11-JAN-1996; 96US-0583787.
 XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
 PA Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
 PI WPI; 1997-372858/34.
 DR P-PSDB; AAMW34558.
 XX New thermostable glycosidase(s) - from Thermococcus, Staphylothermus
 PT and Pyrococcus, used in the textile, food processing,
 PT pharmaceutical, detergent and baking industries
 PS
 PS Claim 4; Fig 5; 82pp; English.
 XX The present sequence encodes glycosidase isolated from Thermotoga
 CC maritima. The enzyme or its encoding nucleic acid sequence is used for
 CC generating glucose from soluble oligosaccharides. The enzyme can be
 CC used in the food processing, pharmaceutical, textile, detergent and
 CC baking industries. The enzyme is also used to treat lactose intolerance
 CC as a diagnostic reporter molecule, in corn wet milling or in the fruit
 CC juice industry. The enzymes can be used to hydrolyse guar gum to remove
 CC non-reducing terminal mannose residues. The nucleic acids encoding the
 CC enzyme may be used to generate probes to identify similar sequences.
 SQ Sequence 2166 BP; 668 A; 454 C; 578 G; 466 T; 0 other.
 Query Match 4.6%; Score 135.8; DB 18; Length 2166;
 Best Local Similarity 53.5%; Pred. No. 4.3e-13;
 Matches 284; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

QY 905 CATGACGGCTCAATGCGATCATGCGCTGTCGACGAGAAACCTAATATCTTGA 964
 DB 618 GATGACGGCTTACAAACAACTGAATGAAATATCTTCAAGAAAGATGCTTTGA 677
 QY 965 TGGAGTCTTCAAAAGAAATGGGGTGGAGTGCCTTAATCATGAGCGACTGACGCAC 1024
 DB 678 GAAGTCTTCAAGGAAAGATGGGATTTGCGGTTTCGATGAGCGACTGCTACGCGG 737
 QY 1025 ATACAGTACCACAGAAAGCCGTGTCGAGCCTCGACCTCGAGATGCCCCG 1075
 DB 738 AGACAACCCCTGTAGAACAGCTCAAGCCGGAACGATATGATATGCTGG 788
 RESULT 14
 AAV36911
 ID AAV36911 standard; DNA; 2166 BP.
 XX AAV36911;
 AC
 XX 21-DEC-1998 (first entry)
 DT
 XX Thermotoga maritima MSB-66 glycosidase gene coding region.
 DE
 XX glycosidase; MSB-66; thermostable enzyme; oligosaccharide;
 KM glucose; sugar; baking; textile; detergent; beta-galactosidase; ss.
 KM Thermotoga maritima strain MSB-66.
 OS
 XX WO9824799-A1.
 PN 11-JUN-1998.
 XX 08-DEC-1997; 97MO-US22623.
 PF
 XX 10-OCT-1997; 97US-0949026.
 PR 06-DEC-1996; 96US-0056916.
 XX (DIVE-) DIVERSA CORP.
 PA Bylina EJ, Lam DE, Mathur EJ, Swanson RV;
 PI WPI; 1998-362407/31.
 DR P-PSDB; AAMW9862.
 XX Glycosidase enzymes from organisms of the genera Staphylothermus,
 PT Pyrococcus and Thermococcus - for deriving sugar from
 PT oligosaccharides, useful in the e.g. food processing, textile or
 FT baking industries
 PS
 PS Claim 1; Fig 5a-b; 92pp; English.
 XX This isolated polynucleotide comprises a coding region for
 CC glycosidase MSB-66 (see AAMW9862) from a Thermotoga maritima MSB8
 CC clone (66) that grows optimally at 85 degC in high salt medium.
 CC The sequence shows 53% nucleic acid identity to beta-galactosidase
 CC B of Clostridium thermocellum. The invention provides 18
 CC polynucleotides (see AAV36907-24) coding for thermostable glycosidases
 CC (see AAMW9858-75) having glycosidase, alpha-galactosidase,
 CC beta-galactosidase, beta-mannosidase, beta-mannanase, endoglucanase
 CC or pullulanase activity. Vectors and host cells are also claimed.
 CC A method is provided for producing the enzymes by recombinant
 CC techniques. A claimed method for generating glucose from soluble
 CC cell oligosaccharides comprises contacting a sample (selected from
 CC dairy products, fruit juice, detergent, textile, guar gum, animal
 CC feed, plant biomass or waste product) containing oligosaccharides
 CC (selected from maltose, cellobiose, lactose, sucrose, raffinose,
 CC starchose, verbasose, cellobiose, starch, amylose, glycogen,
 CC disaccharides, polysaccharides and pullulan) with one of the
 CC claimed glycosidases such that glucose is produced.
 XX
 XX Sequence 2166 BP; 675 A; 444 C; 581 G; 466 T; 0 other;

Query Match 4.6%; Score 135.8; DB 19; Length 2166;
 Best Local Similarity 53.5%; Pred. No. 4.3e-13;
 Matches 284; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

545 GCTCGGTTCCACATTCACCAACACCTCTCGAAGAGCAGCTAAGATGGGCAAGA 604
 258 GCTCGGTTCTACTGAAACAGACGACCTTCTGGAAGAACTGGGAAACCACTGGGAGAGA 317
 605 GGCACATCGCTAAGAGTGGGATGATGATCTCGCCCGGACATCAACAGCCTCCCC 664
 318 ACTTGAAGGAATACGGTGTGATGTCTCTTCGACACCTGCAATCAACAGAAACCC 377
 665 TCTCGGTGACGCTGCTTCGAGTGTGATGAGAGATCCCTCTCGGCGGCTTGGAGC 724
 378 TCTTTGTGGAAGAAATTCGAGTACTACTCAAGATCTGCTCTTCCGGTGAATGCG 437
 725 TCCGCTCTCATCCCGGCAATTCAGAGACGAGGATGAGCTACGATCAAGCACTTTT 784
 438 TTCACCTTTGTCAGAGGATTCATCTCAAGAGGAGGAGCCCTGATTAACACTTTGT 497
 785 GTCAATGATCAGAGAGAGGCGCATGATGTCAGAGCATGCTCAGGAGGCGCTT 844
 438 GCGCAACACAGGAAAGAGAGATGTAGTGTGACACATCTGTCTCAGAGGAGCCCT 557
 845 CCGTAATCTACGCACTCCCTTCAGATTTCTGTGCGAGACTCCAGCCGGGTGCTT 904
 558 CAGAAATAATATCTGAAGGTTTGAATGCTGTCAAGAAAGCAAGACCTTGACCG 617
 905 CATGAGCGCTACATGATGATGATGCTGTGTCAGAGGAAACCTTAATATCTTGA 964
 618 GATGAGCGCTTACAACTGATGATGATGATGATGATGATGATGATGATGATGAT 677
 965 TGGATGCTTGAAGAAATGGGTTGGATGGCTTAATCATGAGGAGCTGTAGCGCAC 1024
 678 GAAAGTTCTCAAGAAATGGGATTTGGCGTTTCGTGATGAGGAGCTGTAGCGCG 737
 1025 ATACATGACCAAGAGCCGTTGGCAGAGCTGACCTGAGATCCCGG 1075
 738 AGACACCTCTGTAGACAGCTCAAGCGCGGAAAGATATGATCATCTCG 788

RESULT 15
 AA199683
 ID AA199683 standard; DNA; 4403765 BP.

AC AA199683;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
 XX
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 variation; epidemiology; patient treatment; epidemic monitoring; ds.
 OS Mycobacterium tuberculosis.
 XX
 PN US6294328-B1.
 XX
 PD 25-SEP-2001.
 XX
 PF 24-JUN-1998; 98US-0103840.
 XX
 PR 24-JUN-1998; 98US-0103840.
 XX
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Felschmann RD, White OR, Fraser CM, Venter JC;
 XX
 DR WPI; 2001-647261/74.
 XX
 PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
 determining the nucleotide sequence of the strain at positions in the
 genome corresponding to positions where M. tuberculosis strains CDC

PT 1551 and H37Rv differ -
 XX
 XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
 XX
 CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
 CC H37Rv (AA199682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
 XX
 SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 3.8%; Score 111.6; DB 22; Length 4403765;
 Best Local Similarity 52.0%; Pred. No. 4.9e-09;
 Matches 345; Conservative 0; Mismatches 289; Indels 30; Gaps 3;

2061 GAGGCGCGCGGACGCGGAGAGATGAGCTCCCGGCGTGTGACACGATGCTGCGCAC 2120
 DB 217690 GAGGCGCTTCCGACAGCGCGGATCTGTGCTCCATGAGGCTGAGATGCTGATCGCGCA 21749
 2121 GTGCGCGCGCGGACCGCAACACCGCTGCTGTCATGACAGAGGCGCACCCCGAGAGATG 2180
 DB 217750 GTGCGCGCGCGGACCGCAACACCGCTGCTGTCATGACAGAGGCGCACCCCGAGAGATG 21809
 2181 CCGGCGCTGACAGCGCGCGCGCGCGCTGATCAGAGCTGTAGCGGCGCACAGAGAGGCG 2240
 DB 217810 CCGGCGCGCTGACAGCGCGCGCGCGCGCTGATCAGAGCTGTAGCGGCGCACAGAGAGGCG 2240
 2241 AACTCATGTCGCGCGCGCGCGCGCGCTGATCAGAGCTGTAGCGGCGCACAGAGAGGCG 2300
 DB 217870 CAGGCGCTTCCGACAGCGCGGATGATGATGATGATGATGATGATGATGATGATGAT 217929
 2301 TTCCCGCAAGCGCGCTGACAGCGCGCGCGCGCTGATCAGAGCTGTAGCGGCGCACAGAGAGGCG 2359
 DB 217930 TTCCCGCGTATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 217989
 2360 -----CACGCTGACGCGCGCGCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2405
 DB 217990 GGGACATCGACACAGATCCACTACACCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 218049
 2406 TTTCGCGACAGAGAGCTCAATTTCCCTTTGGCAGCGCGCTGCTTACACGATTTTGGC 2465
 DB 218050 AGCACAATATGACACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 218109
 2466 TTTTCAATCTGCGCTGAT 2255
 DB 218110 TATGCTGACCTGAT 218166
 2526 AACACCGGCTGCTGCGCGCGCGCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2285
 DB 218167 AACACGGGCGAGCGCGCGCGCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 218226
 2586 GCCAAGATTACCGCGCGCGCGCGCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2645
 DB 218227 GAATTCGCGGTTGCG-----GTTGCGGATTCGAGCGCGGTCGACCTGGAACCC 218274
 2646 GGCAGAGCAGAGCGGCTGAT 2705
 DB 218275 GGCAGACTCGCGGCTGAT 218334
 2706 GAGG 2709
 DB 218335 GAGG 218338

Wed May ' 7 14:14:33 2003

Search completed: May 4, 2003, 00:38:56
job time : 2161.99 secs

us-10-027-000-1.rng

